



University of Pennsylvania  
**ScholarlyCommons**

---

Statistics Papers

Wharton Faculty Research

---

1978

# A Contribution to Kiefer's Theory of Conditional Confidence Procedures

Lawrence D. Brown  
*University of Pennsylvania*

Follow this and additional works at: [http://repository.upenn.edu/statistics\\_papers](http://repository.upenn.edu/statistics_papers)



Part of the [Statistics and Probability Commons](#)

---

## Recommended Citation

Brown, L. D. (1978). A Contribution to Kiefer's Theory of Conditional Confidence Procedures. *Annals of Statistics*, 6 (1), 59-71.  
<http://dx.doi.org/10.1214/aos/1176344065>

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/statistics\\_papers/275](http://repository.upenn.edu/statistics_papers/275)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# A Contribution to Kiefer's Theory of Conditional Confidence Procedures

## **Abstract**

The "procedures" discussed in this paper are of the following type: The statistician makes a conventional decision (in a multiple decision problem). He also provides a statement of the guaranteed conditional probability that his decision will be correct, given the observed value of some conditioning random variable. Various admissibility criteria to relate such procedures are proposed. For example, we say that one procedure is better (first sense) than a second if the guaranteed conditional confidence statement using the first is stochastically larger than that using the second, for all possible states of nature. Some ramifications of these admissibility criteria are discussed, and some specific admissible procedures are described for problems with two possible states of nature. In particular, the procedure having the finest possible monotone conditioning and having equal conditional confidence under both states of nature is shown to have many desirable admissibility properties.

## **Keywords**

conditional confidence procedures, guaranteed conditional confidence

## **Disciplines**

Statistics and Probability

## A CONTRIBUTION TO KIEFER'S THEORY OF CONDITIONAL CONFIDENCE PROCEDURES

BY LAWRENCE D. BROWN<sup>1</sup>

*Rutgers University*

The "procedures" discussed in this paper are of the following type: The statistician makes a conventional decision (in a multiple decision problem). He also provides a statement of the guaranteed conditional probability that his decision will be correct, given the observed value of some conditioning random variable. Various admissibility criteria to relate such procedures are proposed. For example, we say that one procedure is better (first sense) than a second if the guaranteed conditional confidence statement using the first is stochastically larger than that using the second, for all possible states of nature. Some ramifications of these admissibility criteria are discussed, and some specific admissible procedures are described for problems with two possible states of nature. In particular, the procedure having the finest possible monotone conditioning and having equal conditional confidence under both states of nature is shown to have many desirable admissibility properties.

**1. Introduction.** A "conditional confidence procedure" involves a conventional decision procedure and a conditioning random variable. The decision procedures to be considered here are multiple decision (classification) procedures. This is the interpretation given in Kiefer [3]: the statistician makes the conventional decision called for on the basis of the observed random sample. He *also* provides a conditional statement concerning the probability of a correct decision given the observed value of the conditioning random variable, and given each of the possible states of nature. Such procedures are introduced in Kiefer ([2], [3]) and form an essential part of Kiefer's theory.

Both these papers, especially [3], also describe certain partial orderings (notions of betterness) among such conditional decision procedures and discuss the nature of the admissible procedures under these orderings. These partial orderings are based on certain stochastic properties of the conditional confidence statement.

However, the conditional confidence statements described in [3] depend on the true, *but unknown*, state of nature. For many plausible applications such a conditional confidence statement is inappropriate. A statement is required which is not dependent on the unknown state of nature. In this paper we instead suggest as a conditional statement the *guaranteed conditional confidence*, which is the infimum over all states of nature of this conditional confidence (given the value of the conditioning random variable). Following this, some plausible admissibility criteria for such procedures are introduced in Sections 3-5. Example 5.1

---

Received March 1976; revised November 1976.

<sup>1</sup> Research supported in part by NSF Grant MPS-72-05075-A02.

AMS 1970 subject classifications. Primary C05; Secondary A99, C99, F07.

Key words and phrases. Conditional confidence procedures, guaranteed conditional confidence.

presents what we feel is a strong case for the use of one of our criteria in a particular type of classification problem.

One feature shared by our various admissibility criteria is that the admissible procedures are always conditionally maxmin—that is, for each given value of the conditioning variable the procedure maximizes the minimum probability of a correct decision. This feature is discussed in more detail in Section 6. It is described there how this implies that the admissibility criteria operate on the conditioning rules of various procedures, and that certain conditioning rules are better than others. Such a situation is qualitatively different than the one exposed in [3] and [2] for the type of confidence statements and admissibility criteria developed there. (See e.g. [3, Theorem 3.1].) The simple two-parameter point problem is then discussed in the last section of this paper. Some reasonable, admissible procedures are described, corresponding to the different admissibility criteria proposed. This simple problem does have some possible plausible applications. (See the classification example in Section 5.) However these applications alone are not the main justification for the detailed results which appear in this section of the paper. Rather, we feel that these results are a first step toward similar results in more realistic multiple decision settings, involving larger parameter spaces. Also, this careful investigation of such simple problems has deepened our understanding of the criteria proposed in this paper, and we hope the reader will find a similar benefit.

**2. Basic notation.** Consider a statistical problem in abstract form—a measure space  $\mathcal{X}$ ,  $\mathcal{B}$  and a set of possible probability distributions,  $\{F_\omega: \omega \in \Omega\}$ , on  $\mathcal{X}$ ,  $\mathcal{B}$ .

A nonrandomized confidence procedure,  $\delta$ , in Kiefer's theory consists of a decision rule, say  $\beta$ , making decisions in a space  $\mathcal{D}$ ; and a  $\mathcal{B}$ -measurable conditioning random variable,  $Z$ , taking values in  $\mathcal{X}$ ,  $\mathcal{B}_x$  (or its associated  $\sigma$ -field,  $\mathcal{B}_0 \subset \mathcal{B}$ ). [In [2] and [3] the critical regions of the decision rule,  $\beta$ , are often referred to as the " $\mathcal{C}^1$  partition" of the procedure  $\delta$ , and written as  $|\delta|$ . That terminology is not used here.]

Let  $D_\omega$  denote those decisions which are correct for  $\omega$ . For a given procedure consider

$$(2.1) \quad \Gamma_{\omega, \delta}(z) = \Pr_\omega(\beta \in D_\omega | Z = z) .$$

This (measurable) function—which depends also on  $\omega$ —is fundamental in the theory. [In the nonrandomized case discussed here  $\Gamma$  may also be thought of as a function of  $x$ , defined by  $\Gamma(Z(x))$  in the above notation. However, because  $\Gamma$  is written above as a function of  $z$ , certain expressions in Section 3 are valid also in the case of randomized conditioning rules (as described in [3, Section 4]). In particular the optimality assertions and the proofs of Theorems 3.1 and 3.2 are valid within the class of all randomized rules.]

If it were possible, it would be desirable to have  $\Gamma_{\omega, \delta}(z)$  large for all  $\omega$ ,  $z$ . Since this is not possible it is necessary to compare different procedures by

considering some sort of notion of goodness in which “larger”  $\Gamma$ ’s are better (larger in some appropriate sense).

As noted in the introduction the statement  $\Gamma_{\omega,\delta}(z)$  may depend on  $\omega$  for given  $\delta, z$ . It may therefore not be an appropriate summary statement, or estimate, of the statistician’s conditional confidence in the rule  $\beta$ , given  $Z = z$ . A better summary statement is the *guaranteed conditional confidence*, defined by

$$(2.2) \quad \kappa_{\delta}(z) = \inf_{\omega \in \Omega} \Gamma_{\omega,\delta}(z).$$

TECHNICAL COMMENTS. The simple expression in (2.2) ignores two significant technical issues.

First, the heuristics of the situation demand that  $\kappa_{\delta}(z)$  should not be influenced by values of  $\Gamma_{\omega,\delta}(z)$  for which  $\omega$  is a posteriori “impossible” given  $Z = z$ . But since  $\Gamma_{\omega,\delta}$  is defined as a conditional expectation, its value is arbitrary at such values of  $\omega, z$  (subject only to measurability restrictions). An added convention is required to guarantee that  $\Gamma_{\omega,\delta}$  is taken to be large for such impossible values of  $\omega$  so that the infimum on the right of (2.2) is not affected by such values. This convention is described below.

CONVENTION 2.1. Let  $F_{\omega} \ll \nu$  (or, more generally,  $F_{\omega}^Z \ll \nu^Z$ ) for all  $\omega \in \Omega$ , with  $\nu(\nu^Z)$   $\sigma$  finite. Let  $f_{\omega}^Z$  denote some fixed version of  $dF_{\omega}^Z/d\nu^Z$  and let  $S_{\omega} = \{z \in \mathcal{Z} : f_{\omega}^Z > 0\}$ . By convention, always set  $\Gamma_{\omega,\delta}(z) = 1$  if  $z \notin S_{\omega}$ .

This convention, possibly in conjunction with (2.2’), below, will serve to guarantee that  $\kappa$  is essentially uniquely defined a.e. ( $\nu$ ); and that conditionally “impossible” values of  $\omega$  (i.e.  $\{\omega : z \notin S_{\omega}\}$ ) influence  $\kappa$  only on a  $\nu$ -null set.

Note that the dominatedness assumption in the above convention is a fortiori satisfied if either  $\Omega$  or  $\mathcal{Z}$  is countable. We do not know how to proceed in general if  $\Omega$  and  $\mathcal{Z}$  are both uncountable and  $\{F_{\omega}^Z : \omega \in \Omega\}$  is not a dominated family.

Most conditional confidence problems involve finite or countable, decision spaces. In that case (2.2) and Convention 2.1 serve to define  $\kappa_{\delta}$  uniquely a.e. ( $\nu^Z$ ). On the other hand, if the decision space is uncountable, and  $\mathcal{Z}$  is also uncountable, then different determinations of the  $\Gamma_{\omega,\delta}$ ,  $\omega \in \Omega$ , may lead to essentially different determinations of  $\kappa_{\delta}$ . In that case (again assuming that  $F_{\omega}^Z \ll \nu^Z$ ,  $\omega \in \Omega$ ) the most satisfactory course is to replace (2.2) by

$$(2.2') \quad \kappa_{\delta}' = \text{ess inf}_{\omega \in \Omega} \Gamma_{\omega,\delta}$$

where the “ess. inf.” is taken relative to the measure  $\nu^Z$ . (That is to say,  $\kappa_{\delta}'(\cdot) \leq \Gamma_{\omega,\delta}(\cdot)$  a.e.  $\nu^Z \forall \omega$ , and if  $g$  satisfies  $g(\cdot) \leq \Gamma_{\omega,\delta}(\cdot)$  a.e.  $\nu^Z \forall \omega$  then  $g(\cdot) \leq \kappa_{\delta}'(\cdot)$  a.e.  $\nu^Z$ .) This definition, together with Convention 2.1, serves to uniquely define  $\kappa_{\delta}'$  (a.e.  $\nu^Z$ ). Further conditions are required in order that  $\kappa_{\delta}'$  may be computed conditionally from suitable versions of  $\Gamma_{\omega,\delta}$ ,  $\omega \in \Omega$ —that is, in order that  $\kappa_{\delta}' = \kappa_{\delta}$ . Such conditions can often be provided by restricting attention to certain “smooth” (e.g. continuous) versions of the conditional probabilities.

**3. First admissibility criterion.** The first notion of goodness (admissibility criterion) we propose is as follows:

Let

$$(3.1) \quad Q_{\omega, \delta}(r) = \Pr_{\omega}(\kappa_{\delta} \leq r).$$

Then we say that  $\delta_1$  is better than  $\delta_2$  (first sense) if  $\kappa_{\delta_1}$  is stochastically larger than  $\kappa_{\delta_2}$  for every  $\omega$ . In more formal terms this reads— $\delta_2$  is *inadmissible* (first sense) if there is a procedure  $\delta_1$  for which

$$(3.2) \quad Q_{\omega, \delta_1}(r) \leq Q_{\omega, \delta_2}(r) \quad \text{for all } \omega, r$$

with strict inequality for some  $\omega, r$ .

To justify this notion note first that in any experiment the value of  $\omega$  is unknown. Hence  $\kappa_{\delta}$  is the best possible conservative estimate of the conditional probability of making a correct decision, given the value of the conditioning variable,  $Z$ . The experimenter would like to work things out so that his guaranteed conditional confidence will be stochastically large. This leads to the admissibility criterion expressed in (3.2), above.

The formal admissibility criteria in [3] are very different. Nevertheless the discussion there contains a considerable emphasis on describing and discussing procedures in which for each  $x \in \mathcal{X}$   $\Gamma_{\omega}(Z(x))$  is approximately constant as  $\omega$  ranges over  $\Omega$ . See especially [2, Section 1] and [3, Sections 1.1 and 6D]. We believe that this emphasis is due in large part to the feeling that the relevant quantity in many applications is really  $\kappa_{\delta}$  (instead of merely  $\Gamma_{\omega}$ ), for the reason mentioned above. [Note that,  $\kappa_{\delta}(\cdot)$  is symbolically defined (by  $\psi(\cdot)$ ) in [3, Section 1.1].]

Observe that (3.2) is equivalent to

$$(3.3) \quad \int h(r) dQ_{\omega, \delta_1}(r) \geq \int h(r) dQ_{\omega, \delta_2}(r)$$

for all nondecreasing  $h$  and all  $\omega$ , with strict inequality for some  $h, \omega$ .

There is a further relation between the above admissibility criterion and those defined in [3]. Suppose  $\delta_1$  and  $\delta_2$  are two procedures each having  $\Gamma_{\omega, \delta_i}(z)$  constant for each fixed  $z$ . (Hence  $\kappa_{\delta_i} = \Gamma_{\omega, \delta_i}$ ,  $\omega \in \Omega$ .) It is noted above that such procedures are recommended in the discussion in [2] and [3]; and it will be described in Section 6 that our admissibility criteria invariably lead to the use of such procedures. Then, Kiefer's first admissibility criterion (see [3 (2.4)]) states that  $\delta_1$  is better than  $\delta_2$  if and only if our (3.3) holds for all  $h$  of the form  $h(r) = r\chi_{r \geq R}(r)$ ,  $0 \leq R \leq 1$ . For  $\delta_i$  as above, the other admissibility criterion of Kiefer's development (see [3 (2.6)]) is equivalent to:  $\delta_1$  is better than  $\delta_2$  if and only if our (3.3) holds for all  $h$  of the form  $h(r) = r$  for  $r < R$ ,  $= 1$  for  $r \geq R$ . Therefore, among the class of procedures having  $\kappa_{\delta} = \Gamma_{\omega, \delta}$ ,  $\omega \in \Omega$ , our notion of admissibility (first sense) is slightly weaker than the notion referred to in Kiefer's paper [3] as "admissibility in the sense (2.4) (2.6)." See also [3, Section 6D].

**4. Second admissibility criterion.** There is another relevant point raised in [2] and [3]. It is argued that, other features being equal, a large variation in  $\Gamma_{\omega}$  is desirable. (See, for example [3, Section 0].) Because of our previous arguments we would translate this into a principle concerning  $\kappa$ . To place this

argument in more concrete terms consider the following simplified example. Suppose there are two procedures  $\delta_1$  and  $\delta_2$  and that their associated functions satisfy

$$\begin{aligned} Q_{\omega, \delta_2}(r) &= 0 & r < \frac{3}{4} \\ &= 1 & r \geq \frac{3}{4} & \text{for all } \omega \\ Q_{\omega, \delta_1}(r) &= 0 & r < \frac{1}{2} \\ &= \frac{1}{2} & \frac{1}{2} \leq r < 1 \\ &= 1 & r = 1 & \text{for all } \omega. \end{aligned}$$

(Thus, for all  $\omega$ , under  $\delta_2$  the random variable  $\kappa$  is  $\frac{3}{4}$  with probability 1, and under  $\delta_1$  it is  $\frac{1}{2}$  and 1 with equal probability.) Neither procedure would dominate the other in the sense of (3.2). If, in addition, both procedures have constant  $\Gamma_{\omega, \delta}(z)$  for each  $z$  then they both have  $\Pr_{\omega}(\text{correct decision}) = \Pr_{\omega}(D_{\omega}) = \frac{3}{4}$  for all  $\omega$  and so they are equally good unconditionally. However the consideration alluded to above would indicate that  $\delta_1$  is preferable to  $\delta_2$ .

Suppose this consideration is accepted by the experimenter as reflecting his goals for the conditional confidence procedure. Then it indicates as a principle that  $\delta_2$  is *inadmissible (second sense)* if there is another procedure  $\delta_1$  with

$$(4.1) \quad \int h(r) dQ_{\omega, \delta_1}(r) \geq \int h(r) dQ_{\omega, \delta_2}(r)$$

for all nondecreasing, *convex*  $h$  and all  $\omega$ , with strict inequality for some  $\omega$  and some nondecreasing *convex*  $h$ .

**PROPOSITION 4.1.** *Admissibility (second sense) implies admissibility (first sense).*

**PROOF.** Suppose  $\delta_2$  is inadmissible (first sense). Then there exists a  $\delta_1$  such that  $Q_{\omega, \delta_1}(r) \leq Q_{\omega, \delta_2}(r)$  for all  $\omega, r$  with strict inequality for some  $\omega, r$ , say for  $\omega', r'$ . It follows from (3.3) that (4.1) holds for all nondecreasing *convex*  $h$  and all  $\omega$ . It remains to show that there is an appropriate pair  $\omega, h$  such that strict inequality holds in (4.1).

Suppose  $r' = 1$ . Then strict inequality holds in (4.1) for  $\omega = \omega'$  and  $h(t) = \chi_{(1)}(t)$ , which is *convex* on  $[0, 1]$ . If  $r' \neq 1$  then choose  $\omega = \omega'$  and  $h(r) = (r - r')^+$ . Then  $\int_{r'}^1 (1 - Q(r)) dr = \int h(r) dQ(r)$  for any cdf  $Q$  on  $[0, 1]$ . It follows from this (and the right continuity of  $Q$ ) that  $\int h(r) dQ_{\omega', \delta_1}(r) > \int h(r) dQ_{\omega', \delta_2}(r)$ .

**5. Third admissibility criterion.** Both of the above criteria reflect only a vague assessment of goals by the experimenter. When these goals can be refined they result in a third, more precise, criterion. Observe that the function  $h$  which appears in (3.3) and (4.1) plays the role of a gain function. The value of  $h(r)$  can be thought of as representing the "gain", ("satisfaction", etc.) to the experimenter from making a statement with guaranteed conditional confidence  $r$ . Then

$$(5.1) \quad \int h(r) dQ_{\omega, \delta}(r) = E_{\omega}(h(\kappa_{\delta}(z))) = B_h(\omega, \delta)$$

is the expected gain (to be called *benefit*) resulting from the procedure  $\delta$ . The admissibility principles embodied in (3.3) and (4.1) state in this interpretation

that a procedure is inadmissible if another procedure has benefit at least as great for all  $\omega$  no matter what gain function,  $h$ , is used among those in a wide class (increasing functions for (3.3) and increasing convex ones for (4.1)). [This kind of "robustness" in the gain function is analogous to the robustness in the loss function considered in Brown [1] for estimating normal means.]

Suppose, however, the experimenter can settle on a single nondecreasing gain function,  $h$ , to correspond to his goals in performing the experiment. Then his ranking of procedures takes on the traditional decision theoretic structure: a procedure  $\delta_2$  is *inadmissible relative to a given gain function,  $h$* , if there is another procedure  $\delta_1$  such that

$$(5.2) \quad B_h(\omega, \delta_1) \geq B_h(\omega, \delta_2) \quad \text{for all } \omega \in \Omega$$

with strict inequality for some  $\omega$ .

[The structure here is "traditional" in the sense of being based on a single real-valued function,  $B$ . It is of course really the negative of the usual theory in that it is based on gain and benefit instead of loss and risk.]

It is now also possible to describe a property stronger than second sense admissibility. Suppose that for each nondecreasing convex  $h$ ,  $\delta$  is admissible relative to  $h$ . Then we say that  $\delta$  is *totally admissible* (second sense). ("Total admissibility (first sense)" could be similarly defined, but the definition appears to be barren; in certain simple examples where we have investigated this property it appears that there do not exist any nontrivial totally admissible (first sense) procedures.)

Minimax (or, rather, maxmin) notions can also be defined. Two useful notions are as follows:  $\delta_1$  is called *maxmin relative to  $h$*  if

$$(5.3) \quad \min_{\omega \in \Omega} B_h(\omega, \delta_1) = \max_{\delta} \min_{\omega \in \Omega} B_h(\omega, \delta).$$

$\delta_1$  is called *totally maxmin* (second sense) if it is maxmin relative to  $h$  for each nondecreasing convex  $h$ . Theorem 7.2 describes a totally admissible and totally maxmin procedure (second sense) for a special type of simple confidence problem.

The considerations alluded to in discussing the second sense of admissibility indicate that in many situations an experimenter who settles on a single gain function,  $h$ , will choose  $h$  to be convex. An intuitively appealing choice for  $h$  is  $h(r) = r^2$ . Under this choice of  $h$ , if two procedures have the same expected guaranteed confidence unconditionally—that is if

$$\int r dQ_{\omega, \delta_1}(r) = \int r dQ_{\omega, \delta_2}(r)$$

for all  $\omega$ —then the procedure with the larger variance of the guaranteed conditional confidence for all  $\omega$  (if there is such a procedure) will be the better procedure. [These considerations might also motivate one to adopt a variant of the above admissibility criterion in certain situations, and to base admissibility on only the two benefit functions  $B_{h_1}$  and  $B_{h_2}$  where  $h_i = r^i$ .]

However, convex gain functions are not always appropriate. Consider the following example in which a nonconvex gain function is clearly called for.



EXAMPLE 5.1. Consider a simple (nonBayesian) classification situation in which two types of items are produced. They are labelled by  $\omega = 0$  and  $\omega = 1$ . A given item cannot be directly tested to check which type it is but some supplementary related characteristic(s),  $X$ , can be measured.  $X$  has known distribution  $F_\omega$ . Items which can be classified as  $\omega = 0$  or  $\omega = 1$  with confidence,  $c$ , say, may be sold as "classified items" at a (premium) per item price,  $p$ . Those items which cannot be so classified must be sold in mixed lots at a lower per item price, say  $q$ . It follows that the appropriate gain function to the producer of these items is

$$\begin{aligned} h(r) &= q \quad r < c \\ &= p \quad r \geq c. \end{aligned}$$

There exists an optimal conditional confidence procedure for such a gain function. It is described in Theorem 7.1.

[Note the symmetry in the above formulation. As given the formulation guarantees, for example, that if a buyer purchases only those items which are offered at the premium price and if  $\omega = i$  then at least  $100c\%$  of such items will be classified as  $\omega = i$ ; for both  $i = 0, 1$ . One might instead wish to require that under these two conditions a proportion  $c_i$  of such items will be classified as of type  $i$ , with  $c_0 \neq c_1$ . Although the formulation of confidence problems which is described in this note is specifically symmetrical, beginning with the definition (2.2) of  $\kappa$ , this asymmetrical situation can be formulated and discussed in a manner similar to the considerations here.]

**6. Importance of the conditioning rule.** One surprising (and disappointing) consequence of the notions of admissibility adopted in [3] is that admissibility of a procedure rarely, if ever, depends on the choice of conditioning variable (or subfield). In fact, in any two-parameter problem a procedure is admissible if and only if the decision rule  $\beta$  is unconditionally admissible, irrespective of the conditioning variable used. (See especially [3, Theorem 3.1 and Corollary 5.3] or [2, Theorem 3.1].)

The situation is diametrically different for the criteria we have suggested. Here, for each conditioning random variable,  $Z$ , there is usually a best decision rule—say  $\beta_Z$ . Since the value(s) of  $B_h$ , on the basis of which admissibility is judged, depend on  $(\beta_Z, Z)$ , they are ultimately a function only of the conditioning rule described by  $Z$ . These admissibility criteria therefore operate in a non-trivial way on conditioning rules (together with their associated "best" decision rules). The construction of this "best" rule  $\beta_Z$  is sketched below.

Assume the existence for each  $\omega \in \Omega$  of a regular conditional distribution on  $\mathcal{X}$  given  $Z$ . Thus for each value of  $Z$ ,  $\{F_\omega(\cdot | z) : \omega \in \Omega\}$  is a family of probability distributions on  $\mathcal{X}$ ,  $\mathcal{B}$ . Now, for each fixed  $z \in \mathcal{X}$  consider the statistical decision problem with these distributions  $F_\omega(\cdot | z)$  and with gain function  $G(d, \omega) = \chi_{D_\omega}(d)$ . (Actually, if the  $F_\omega$  do not have identical supports then  $G(d, \omega)$  should be defined as  $\chi_{D_\omega}(d)$  if  $z \in S_\omega$  and 1 if  $z \notin S_\omega$ . See the Technical Comments

in Section 2.) Let  $\beta(\cdot | z)$  denote the maxmin procedure for this conditional problem.  $\beta(\cdot | z)$  may be randomized but that possibility requires no special handling in what follows. (If  $\{F_\omega(\cdot | z) : \omega \in \Omega\}$  is a dominated family then the existence of  $\beta(\cdot | z)$  is guaranteed by results in Le Cam [4].) (If  $F_\omega \ll \nu$ ,  $\omega \in \Omega$  and the regular conditional distribution of  $\nu$  exists, then  $F_\omega(\cdot | \cdot)$  may be taken so that  $F_\omega(\cdot | z) \ll \nu(\cdot | z)$  for all  $\omega \in \Omega$ ,  $z \in \mathcal{X}$ .)

Define  $\beta_Z(x) = \beta(x | Z(x))$ , assuming it is a measurable decision rule. The following is the desired result.

**PROPOSITION.** Assume  $\beta_Z$  exists. Let  $\delta_Z = (\beta_Z, Z)$  and let  $\delta'$  be any other procedure with the same conditioning random variable  $Z$ . Then  $Q_{\omega, \delta_Z}(r) \leq Q_{\omega, \delta'}(r)$  for all  $\omega, r$ . It follows that  $\delta_Z$  is at least as good as  $\delta'$  in any of the three senses of admissibility described above.

**PROOF.** Consider any procedure  $\delta'(\beta', Z)$  with conditioning random variable  $Z$ . Then  $\Gamma_\omega$  is defined by  $\Gamma_\omega(z) = \Pr\{\delta \in D_\omega | z\} = E(G(\beta, \omega) | z)$ . Hence  $\kappa(z)$  is maximized by choosing  $\beta$  at  $Z(x) = z$  so as to maximize  $\min_{\omega \in \Omega} E_\omega(G(\beta, \omega) | z)$ . Since  $\beta_Z$  does precisely this, it follows that  $\kappa_{\omega, \delta_Z}(Z(x)) \geq \kappa_{\omega, \delta'}(Z(x))$  for all  $x$ . The claimed relation between the cumulative distribution functions of  $\kappa_{\omega, \delta_Z}$  and of  $\kappa_{\omega, \delta'}$  follows immediately.

**7. Admissibility in two-point parameter spaces.** This section contains some theorems and examples involving the simplest type of conditional confidence problem—for a two point parameter space in which  $\Omega = \{0, 1\}$ ,  $D = \{0, 1\}$ , and  $D_\omega = \omega$ . In such a situation let  $F_\omega \ll \nu$  and  $f_\omega = dF_\omega/d\nu$ ;  $\omega = 0, 1$ . [Some of the following results require the use of truly randomized procedures. But these procedures all have a relatively simple structure so we have not felt it necessary to introduce the general terminology from [3, Section 4].]

The first result concerns the situation described in Example 5.1. It describes the admissible (third sense) procedures relative to a given gain function of the form  $h(r) = \chi_{\{r \geq c\}}$ . (Note that here  $B_\omega(h, \delta) = P_\omega[\kappa_\delta(Z) \geq c]$ .)

**THEOREM 7.1.** Let  $h = \chi_{\{r \geq c\}}$ ,  $c > \frac{1}{2}$ . Let  $\delta = (\beta, Z)$  satisfy

$$\begin{aligned} Z &= 1 && \text{if } f_0/f_1 < k_1 \text{ or } f_0/f_1 > k_2; \\ (1) \quad &= 1 && \text{with probability } \gamma_i \text{ if } f_0/f_1 = k_i, \quad i = 1, 2; \\ &= 0 && \text{otherwise} \end{aligned}$$

where  $k_1 < k_2$ , and

$$\begin{aligned} \beta &= 1 && \text{if } f_0/f_1 \leq k_1 \text{ and } Z = 1; \\ (2) \quad &= 0 && \text{if } f_0/f_1 \geq k_2 \text{ and } Z = 1; \\ &= \text{anything in } \{0, 1\} && \text{if } Z = 0 \end{aligned}$$

where  $k_i, \gamma_i$  are determined by the conditions

$$(3) \quad \Pr_\omega(\beta = \omega, Z = 1) / \Pr_\omega(\beta \neq \omega, Z = 1) = c / (1 - c), \quad \omega = 0, 1.$$

If no solution to (1), (2), (3) exists then let  $\delta$  be a solution to (3) satisfying

$$(4) \quad \begin{aligned} Z &= 1 && \text{if } f_0/f_1 \neq k; \\ &= 1 && \text{with probability } \gamma_1 \text{ if } f_0/f_1 = k; \\ &= 0 && \text{otherwise} \end{aligned}$$

and

$$(5) \quad \begin{aligned} \beta &= 1 && \text{if } f_0/f_1 > k \text{ and } Z = 1; \\ &= 1 && \text{with probability } \gamma_2 \text{ if } f_0/f_1 = k \text{ and } Z = 1; \\ &= 0 && \text{otherwise if } Z = 1; \\ &= \text{anything in } \{0, 1\} && \text{if } Z = 0. \end{aligned}$$

If no solution to either of the above sets of equations exists, let  $Z \equiv 1$  and let  $\beta$  define the unconditional minimax procedure. Then  $\delta$ , as defined above, is an optimal procedure.

REMARK. A uniqueness assertion can be added to the above theorem. Except in the third case above, on the set on which  $Z = 1$ ,  $\delta$  is essentially (a.e.  $F_0 + F_1$ ) uniquely defined as a function of the sufficient statistic  $f_0/f_1$ . In the third case above there can be other choices of  $Z$  which also yield optimal procedures and, in general, on the set where  $Z = 0$  a.e. ( $F_0 + F_1$ ) the choice of procedure is irrelevant.

PROOF. Let  $\delta' = (\beta', Z)$  be any conditional confidence procedure. Let  $U = \{z: \kappa_{\delta'}(z) \geq c\}$ . Consider the new procedure  $\delta'' = (\beta', Z'')$  where  $Z'' = \chi_U(Z)$ . Then

$$\Pr_{\omega}(\beta' \in D_{\omega} | Z'' = 1) \geq \inf_{z \in U} \Pr_{\omega}(\beta' \in D_{\omega} | Z = z) \geq c.$$

It follows via the definitions that  $\delta''$  is at least as good as  $\delta'$ .

The above shows that every procedure is dominated by some other procedure with  $\text{Range } Z = \{0, 1\}$ . Let  $\delta' = (\beta', Z'')$  be a procedure of this type. If both  $\kappa_{\delta'}(0) \geq c$  and  $\kappa_{\delta'}(1) \geq c$  then  $B_h(\omega, \delta') = 1$ ,  $\omega = 0, 1$ . The fact that  $\kappa_{\delta'}(z) \geq c$ ,  $z = 0, 1$ , implies that  $\Pr_{\omega}(\beta = \omega) \geq c$  for both  $\omega = 0, 1$ . Hence the procedure with  $Z \equiv 1$  and with  $\beta$  the (unconditional) minimax rule is exactly as good as  $\delta'$ . This latter rule is of the third form, above.

Now suppose (w.l.o.g.) that  $\kappa_{\delta'}(0) < c$  and  $\kappa_{\delta'}(1) \geq c$ . Then  $B_h(\omega, \delta') = \Pr_{\omega}(Z'' = 1) \leq c^{-1} \Pr_{\omega}(\beta' = \omega, Z'' = 1)$  since  $\kappa_{\delta'}(1) \geq c$ ;  $\omega = 0, 1$ . Consider the problem of maximizing  $\Pr_0(\beta = 0, Z = 1)$  subject to the side conditions  $\kappa_{\delta'}(1) \geq c$ , that is

$$(1 - c) \Pr_0(\beta = 0, Z = 1) \geq c \Pr_0(\beta = 1, Z = 1),$$

and

$$(1 - c) \Pr_1(\beta = 1, Z = 1) \geq c \Pr_1(\beta = 0, Z = 1).$$

This is a standard decision theoretic problem, covered by Lehmann [5, Chapter 3, Theorem 5]. The maximizing procedure is any procedure of the form,  $\delta$ , described in the statement of the theorem. This procedure also satisfies  $\Pr_0(Z = 1) = c^{-1} \Pr_0(\beta = 0, Z = 1)$  unless  $\Pr_0(Z = 1) = \Pr_1(Z = 1) = 1$ . Hence this

procedure also maximizes  $\Pr_0(Z = 1)$  subject to these side conditions. Symmetrically, the same procedure also maximizes  $\Pr_1(Z = 1)$  subject to the side conditions. This proves the theorem. The uniqueness assertions given in the remark may also be deduced from the above considerations.

**SYMMETRIC PROBLEMS.** The next result concerns second-sense admissibility in the "symmetric" situation. The general definition of the "symmetric" case, which is treated at length in [2], is that the distribution under  $F_0$  of the random variable  $f_1/f_0$  should be the same as the distribution under  $F_1$  of the random variable  $f_0/f_1$ . One of the simplest such situations is where  $\chi \sim N(\mu_\omega, \sigma^2)$ ,  $\omega = 0, 1$  with  $\mu_0 \neq \mu_1$ .

An intuitively appealing procedure in this case is that corresponding to the *maximal symmetric monotone conditioning*. See [2, Section 3.5]. This procedure,  $\delta_0$ , can be described as follows:

$$\begin{aligned} Z_0 &= \max(f_0, f_1)/(f_0 + f_1); \\ \beta_0 &= 1 \quad \text{if } f_0/f_1 < 1; \\ &= 1 \quad \text{with probability } \frac{1}{2} \quad \text{if } f_0/f_1 = 1; \\ &= 0 \quad \text{if } f_0/f_1 > 1. \end{aligned}$$

There is a generalization of this procedure to asymmetric problems. This is described above Theorem 7.4.

**THEOREM 7.2.** *Consider the symmetric case and the procedure corresponding to the maximal symmetric monotone conditioning. Then,  $\delta$  is totally maxmin and totally admissible (second sense).*

**REMARKS.** Again a uniqueness assertion is possible. If the distribution of  $f_1/f_0$  is nonatomic then the above procedure is the essentially unique (a.e.  $F_0 + F_1$ ) procedure having the given properties and depending only on the sufficient statistic  $f_1/f_0$ .

**PROOF.** Let  $0 < c < 1$  and let  $h(r) = (r - c)^+$ . The proof begins with a demonstration that  $\delta_0$  is "Bayes" relative to  $h$  and the symmetric prior, which places equal probability on  $\omega = 0, 1$ .

For any procedure  $\delta$  let  $S_\delta = \{z: \kappa_\delta(z) \geq c\}$ . Then the expected benefit of  $\delta$  under the symmetric prior is

$$\begin{aligned} &B(0, \delta)/2 + B(1, \delta)/2 \\ &= E_0((\kappa_\delta - c)^+)/2 + E_1((\kappa_\delta - c)^+)/2 \\ (7.1) \quad &\leq E_0((\Pr_0(\beta = 0 | Z) - c)\chi_{S_\delta})/2 + E_1((\Pr_1(\beta = 1 | Z) - c)\chi_{S_\delta})/2 \\ &= [\Pr_0(\beta = 0, S_\delta) + \Pr_1(\beta = 1, S_\delta) - c(\Pr_0(S_\delta) + \Pr_1(S_\delta))]/2 \\ &\leq (\frac{1}{2}) \int_{T_\delta} [\max(f_0, f_1) - c(f_0 + f_1)] d\nu, \end{aligned}$$

where  $T_\delta = Z^{-1}(S_\delta)$ .

The right hand side of (7.1) is maximized over arbitrary sets,  $T$ , by choosing  $T = \{x: \max(f_0, f_1) \geq c(f_0 + f_1)\} = \{x: Z_0(x) \geq c\}$ .

Compute that  $\kappa_\delta(Z_0(x)) = Z_0(x)$  in view of the symmetry of the problem and of  $(\beta_0, Z_0)$ . See for example [2, Section 1 (following (1.3)) and Section 3.5]. Hence  $T_{\delta_0} = \{x: Z_0(x) \geq c\}$ .

Furthermore  $\Pr_0(\beta_0 = 0 | Z_0) = \Pr_0(\beta_0 = 1 | Z_0) = \kappa_{\delta_0}$  and  $\beta_0 = i: f_i = \max(f_0, f_1)$ . Hence equality holds throughout (7.1) for the procedure  $\delta_0$ . This shows that  $\delta_0$  is Bayes relative to  $h(r) = (r - c)^+$  and the symmetric prior.

If  $h = \chi_{(1)}$  then  $B_h(\omega, \delta) \leq \Pr_\omega(f_0/f_1 = 0 \text{ or } \infty) = B_h(\omega, \delta_0)$  for  $\omega = 0, 1$ . Hence  $\delta_0$  is optimum for this particular  $h$ , and so a fortiori is Bayes for the symmetric prior.

For any given nondecreasing convex  $h$  on  $[0, 1]$  and any  $\varepsilon \geq 0$  it is possible to write  $0 \leq h - \sum_{i=1}^n a_i h_i \leq \varepsilon$  where each  $a_i \geq 0$  and each  $h_i$  is either of the form  $\chi_{(1)}$  or  $(r - c_i)^+$ . By the above,  $\delta_0$  is Bayes relative to  $\sum_{i=1}^n a_i h_i$  and the symmetric prior. Since  $B(h, \delta) \leq h(1)$  for all  $\delta$  this yields that  $\delta_0$  is  $\varepsilon h(1)$ -Bayes relative to the symmetric prior. Hence  $\delta_0$  is Bayes, since  $\varepsilon$  is arbitrary in the above.

Since  $\delta_0$  is Bayes relative to  $h$  and the symmetric prior and since  $B_h(0, \delta_0) = B_h(1, \delta_0)$  (by symmetry) it follows that  $\delta_0$  is maxmin and admissible relative to  $h$ . Hence  $\delta_0$  is totally maxmin and totally admissible.

It is tempting to believe that  $\delta_0$ , as defined above, is actually an optimum (second sense) procedure. In other words to believe that  $B_h(\omega, \delta_0) \leq B_h(\omega, \delta)$  for all convex nondecreasing  $h$ , all  $\omega \in \Omega$ , and all  $\delta$ . This proposition appears to us to be true if the alternative procedures,  $\delta$ , are restricted to be monotone. However it is not, in general, true. The following simple example shows that  $\delta_0$  is not optimum in the simple symmetric situation described there. Careful perusal of the example should suffice to convince the reader that for most symmetric problems and for many reasonable gain functions  $\delta_0$  is not optimum. (One prominent exception would be the very special case where  $f_0/f_1$  takes on just two values (a.e.  $F_0 + F_1$ ).)

EXAMPLE 7.3 ( $\delta_0$  is not optimum (second sense)). Let  $f_0(x) = [x]/6$  for  $1 \leq x \leq 4$  where "[ ]" denotes "largest integer in". Let  $f_1 = \frac{2}{3} - f_0$ . This problem is symmetric. Let  $h(r) = r$ . Choose constants  $a, b$  which satisfy

$$(7.2) \quad 3a^2 = b^2 - 4b + \frac{1}{4}^5$$

$$(7.3) \quad (1 - a + 2b)(3(1 - a) + 2b) = b^2 - 4b + \frac{1}{4}^5$$

and  $\frac{1}{2} < b < 1, 0 < a < 1$ . The following reasoning shows that a solution exists to (7.2), (7.3). For every  $b \in (\frac{1}{2}, 1)$ , (7.2) always has one solution  $a \in (0, 1)$ . If  $b = \frac{1}{2}$  and (7.2) is satisfied with  $0 < a < 1$  (i.e.  $a = +(\frac{2}{3})^{\frac{1}{2}}$ ) then the left side of (7.3) is smaller than the right, but if  $b = 1$  and (7.2) is satisfied with  $0 < a < 1$  (i.e.  $a = \frac{1}{2}$ ) then the opposite inequality relates the two sides of (7.3). Hence, continuity implies that a suitable solution exists to this pair of equations.

Now, consider the procedure  $\delta_1$ , say, defined by

$$\begin{aligned}\beta_1, Z_1 &= (1, 1) \quad \text{if } 1 \leq x < a; \\ &= (0, 1) \quad \text{if } 2 + b \leq x < 3 - (1 - b)/2 \quad \text{or } \frac{7}{2} \leq x < 4; \\ &= (1, 2) \quad \text{if } a \leq x < 2 + b; \quad \text{and} \\ &= (0, 2) \quad \text{if } 3 - (1 - b)/2 \leq x < \frac{7}{2}.\end{aligned}$$

The procedure  $\delta_1$  is not monotone. It has  $\text{Range } Z = \{1, 2\}$ . It is straightforward to check, using (7.2), (7.3), that  $\Pr_0(\beta_1 = 0 \mid Z_1 = i) = \Pr_1(\beta_1 = 1 \mid Z_1 = i)$ ,  $i = 1, 2$ . Hence  $B(1, \delta_1) = E_1(h(\kappa_{\delta_1}(Z(x)))) = E_1(\Pr_1(\beta_1 = 1 \mid Z_1)) = \Pr_1(\beta_1 = 1) = \frac{1}{2} + b/3 > \frac{2}{3}$ . Meanwhile  $B(1, \delta_0) = \Pr_1(\beta_0 = 1) = \frac{2}{3}$ . Hence  $\delta_0$  does not dominate  $\delta_1$ . (On the other hand  $B(0, \delta_1) = \frac{5}{8} - \frac{6}{8} < \frac{2}{3} = B(0, \delta_0)$  so that  $\delta_1$  does not dominate  $\delta_0$ , so this example is consistent with Theorem 7.2.)

**THE ASYMMETRIC CASE.** The situation in the asymmetric case is not as transparent. The appropriate generalization of the procedure  $\delta_0$  of Theorem 7.2 is the *maximal monotone procedure* having  $E_0(\beta = 0 \mid Z) = E_1(\beta = 1 \mid Z)$ . This procedure is fully described in [2, Section 3.5]. Briefly, it can be obtained by first rewriting the problem via a 1-1 change of variables so that  $f_0/f_1$  is a nondecreasing function, then setting  $Z_0(x) = \min(\int_{-\infty}^x (f_0(t)f_1(t))^{1/2} \nu(dt), \int_x^{\infty} (f_0(t)f_1(t))^{1/2} \nu(dt))$ , and then choosing  $\beta_0 = 1$  or 0 according to whether the first or second of the preceding integrals is larger. (Take  $\beta_0 = 1$  with probability  $\frac{1}{2}$  if the two integrals are equal.)

In this form a special role is played by the value  $c_0$  such that  $\int_{c_0}^{\infty} f_0(t)f_1(t)\nu(dt) \leq \frac{1}{2}$  and  $\int_{c_0}^{\infty} f_0(t)f_1(t)\nu(dt) \leq \frac{1}{2}$ , since  $\beta$  decides 1 or 0 if  $x < c_0$  or  $x > c_0$ , respectively. (If  $\nu(\{c_0\}) > 0$  then  $\beta$  will be randomized at  $c_0$ , in order to guarantee  $E_0(\beta = 0 \mid Z) = E_1(\beta = 1 \mid Z)$ .)

If  $h(r) = r$  then  $B_h(\omega, \delta) = \Pr_{\omega}(\beta = \omega)$ . It is easy to construct many asymmetric examples in which  $\nu(\{c_0\}) = 0$  to avoid randomization and  $\Pr_1(x < c_0) \neq \Pr_0(x > c_0)$ . In any such example  $\delta_0$  will not be maxmin relative to  $h(r) = r$ . In general, no matter what the given  $h$ ,  $\delta_0$  may not be maxmin relative to  $h$ . Of course this is more than enough to see that  $\delta_0$  will not in general be totally maxmin.

However, it is reasonable to conjecture that  $\delta_0$  is totally admissible (second sense). The following theorem demonstrates a fairly strong admissibility property of  $\delta_0$ . The property is interesting on its own merit and motivates the conjecture that  $\delta_0$  is totally admissible.

**THEOREM 7.4.** *Let  $h(r) = (r - c)^+$ . Then the maximal monotone procedure is admissible relative to  $h$ .*

**PROOF.** Fix  $c \geq \frac{1}{2}$ . (The values  $c < \frac{1}{2}$  are uninteresting.) It will be shown that  $\delta_0$  is Bayes relative to a prior which gives some mass  $p = p(c)$  to  $\omega = 0$ ,  $0 < p < 1$ . The admissibility assertion follows immediately from this fact. [Note in the following argument that  $p$  depends partly on the choice of  $c$ . For

this reason the Bayes property described above cannot be used to prove total admissibility as in the symmetric case of Theorem 7.2.]

For any procedure  $\delta$  let  $T_\delta = \{x: \kappa(Z(x)) \geq c\}$ . (The argument requires trivial modifications if  $Z$  is randomized.) Then

$$(7.4) \quad \begin{aligned} & pB_h(0, \delta) + qB_h(1, \delta) \\ & \leq p[\Pr_0(T_\delta, \beta = 0) - c \Pr_0(T_\delta)] + q[\Pr_1(T_\delta, \beta = 1) - c \Pr_1(T_\delta)] \\ & = pE_0((1 - c)\varphi_0 - c\varphi_1) + qE_1(-c\varphi_0 + (1 - c)\varphi_1) \end{aligned}$$

where  $\varphi_i = \chi_{T_\delta, \beta=i}$   $i = 1, 2$ . The right side of (7.4) is maximized by choosing

$$\varphi_0 = \chi_{\{x: f_0/f_1 \geq cq/p(1-c)\}}$$

$$\varphi_1 = \chi_{\{x: f_0/f_1 \leq (1-c)q/cp\}}.$$

(Actually  $\varphi_0$  and  $\varphi_1$  are irrelevant at any  $x$  such that  $f_0(x)/f_1(x) = cq/p(1 - c) = (1 - c)q/cp$ , which can occur if, and only if,  $c = \frac{1}{2}$ .)

Suppose  $z = Z_0(a) = Z_0(b)$  with  $a < c_0 < b$ . Computations then yield that

$$(7.5) \quad \kappa(z) = f_0(a)/(f_0(a) + \Delta f_0(b)) + \Delta f_1(b)/(f_1(a) + \Delta f_1(b))$$

where  $\Delta = (f_0(a)f_1(a)/f_0(b)f_1(b))^{\frac{1}{2}}$ . Let  $a_c$  denote the ( $a$ ) value of  $a \leq c_0$  such that  $\kappa(z_c^-) \geq c$  and  $\kappa(z_c^+) \leq c$  where  $z_c = Z_0(a_c)$  and let  $b_c$  be the corresponding value of  $b \geq c_0$ . The above expression for  $\kappa(z)$  yields that

$$(7.6) \quad f_0(a_c)/f_1(a_c) = (\kappa(z_c)/(1 - \kappa(z_c)))^2(f_0(b_c)/f_1(b_c)).$$

Now, choose  $p$  so that  $f_0(a_c^-)/f_1(a_c^-) \leq (1 - c)q/cp$  and  $f_0(a_c^+)/f_1(a_c^+) \geq (1 - c)q/cp$ . By construction  $f_0(x)/f_1(x) \leq (1 - c)q/cp$  if and only if  $x \leq a_c$  ( $x < a_c$ ) if  $\kappa(z_c) \geq c$  ( $\kappa(z_c) < c$ , respectively). This is precisely the set on which  $\kappa(z) \geq c$  and  $\beta = 1$ .

It follows algebraically from the above equations that  $f_0(x)/f_1(x) \geq cq/p(1 - c)$  if and only if  $x \geq b_c$  ( $x > b_c$ ) if  $\kappa(z_c) \geq c$  ( $\kappa(z_c) < c$ , respectively). This is the set on which  $\kappa(z) \geq c$  and  $\beta = 0$ .

The above facts show that  $\delta_0$  maximizes  $pB_h(0, \delta_0) + qB_h(1, \delta_0)$  so that  $\delta_0$  is Bayes, as desired. This completes the proof.

#### REFERENCES

- [1] BROWN, L. D. (1975). Estimation with incompletely specified loss functions (the case of several location parameters). *J. Amer. Statist. Assoc.* **70** 417-427.
- [2] KIEFER, J. (1975). Conditional confidence statements. *J. Amer. Statist. Assoc.* (To appear).
- [3] KIEFER, J. (1976). Admissibility of conditional confidence procedures. *Ann. Statist.* **4** 836-865.
- [4] LE CAM, L. (1955). An extension of Wald's theory of statistical decision functions. *Ann. Math. Statist.* **26** 69-81.
- [5] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. J. Wiley, New York.

DEPARTMENT OF STATISTICS  
RUTGERS UNIVERSITY  
NEW BRUNSWICK, N.J. 08903